A theory of neural dimensionality and dynamics

In a wide variety of experimental paradigms, neuroscientists tightly control behavior, record many trials, and obtain trial averaged neuronal firing rate data from hundreds of neurons, in circuits containing millions to billions of behaviorally relevant neurons. Such datasets are often analyzed by dimensionality reduction methods that allow us to visualize neuronal dynamics through their projections onto a number of basis patterns. Strikingly, recordings from hundreds of neurons can often be described using a much smaller number of dimensions (basis patterns), and the resulting projections yield a remarkably insightful dynamical portrait of neural circuit computation. Thus many neuronal datasets are surprisingly simple, and we seem to be able to extract reasonable collective neuronal dynamics despite overwhelming levels of neuronal subsampling. This ubiquitous simplicity raises several profound and timely conceptual questions. What is the origin of this simplicity? What does it tell us about the complexity of brain dynamics? Would neuronal datasets become more complex if we recorded more neurons? How and when can we trust dynamical portraits obtained from only hundreds of neurons in a circuit containing billions of neurons? More generally, what, if anything, can we learn about a complex dynamical system by measuring an infinitesimal fraction of its degrees of freedom? We present a theory of neural dimensionality and dynamics that answers these questions, and we further test this theory in neural recordings from monkeys performing reaching movements.